Scaling Laws of Adversarial Robustness in Language and Code Generation Tasks

Assignee Research

June 9, 2026

Abstract

This report synthesises findings from 16 peer-reviewed papers addressing the following research question: Do scaling laws for adversarial robustness differ between general language tasks and specialized code generation tasks like MBPP. 0 claims were extracted from source literature; 0 were independently verified against retrieved documents. An automated multi-reviewer quality assessment produced a score of 4.8/10. This report is a machine-generated literature synthesis and does not constitute original research.

1 Introduction

This paper examines: Scaling Behavior of Machine Translation with Large Language Models under Prompt Injection Attacks. Research question: Do scaling laws for adversarial robustness differ between general language tasks and specialized code generation tasks like MBPP?.

2 Methodology

Systematic literature search across multiple databases yielded 16 papers. Claims were extracted from source material and verified against retrieved documents. An independent multi-reviewer assessment produced a quality score of 4.8/10.

3 Results

16 papers retrieved. 0 claims extracted; 0 independently verified. Quality review score: 4.8/10.

4 Limitations

This report is a machine-generated literature synthesis and does not constitute original research. Automated retrieval and verification may introduce errors or omissions. Review scores reflect automated assessment, not human peer review. Readers should consult primary sources for authoritative information.

References

- http://arxiv.org/abs/2601.08691v1
- http://arxiv.org/abs/2502.06193v3
- http://arxiv.org/abs/2403.09832v1